A Contraction Theory Approach to Stochastic Incremental Stability

Quang-Cuong Pham*
LPPA, Collège de France
Paris, France
cuong.pham@normalesup.org

Nicolas Tabareau
LPPA, Collège de France
Paris, France
tabareau.nicolas@gmail.com

Jean-Jacques Slotine
Nonlinear Systems Laboratory, MIT
Cambridge, MA 02139, USA
jjs@mit.edu

February 1, 2008

Abstract

We investigate the incremental stability properties of Itô stochastic dynamical systems. Specifically, we derive a stochastic version of nonlinear contraction theory that provides a bound on the mean square distance between any two trajectories of a stochastically contracting system. This bound can be expressed as a function of the noise intensity and the contraction rate of the noise-free system. We illustrate these results in the contexts of stochastic nonlinear observers design and stochastic synchronization.

1 Introduction

Nonlinear stability properties are often considered with respect to an equilibrium point or to a nominal system trajectory (see e.g. [31]). By contrast, *incremental* stability is concerned with the behaviour of system trajectories *with respect to each other*. From the triangle inequality, global exponential incremental stability (any two trajectories tend to each other exponentially) is a stronger property than global exponential convergence to a single trajectory.

Historically, work on deterministic incremental stability can be traced back to the 1950's [23, 7, 16] (see e.g. [26, 20] for a more extensive list and historical discussion of related references). More recently, and largely independently of these earlier studies, a number of works have put incremental

*To whom correspondance should be addressed.

stability on a broader theoretical basis and made relations with more traditional stability approaches [14, 32, 24, 2, 6]. Furthermore, it was shown that incremental stability is especially relevant in the study of such problems as state detection [2], observer design or synchronization analysis.

While the above references are mostly concerned with *deterministic* stability notions, stability theory has also been extended to *stochastic* dynamical systems, see for instance [22, 17]. This includes important recent developments in Lyapunov-like approaches [12, 27], as well as applications to standard problems in systems and control [13, 34, 8]. However, stochastic versions of incremental stability have not yet been systematically investigated.

The goal of this paper is to extend some concepts and results in incremental stability to stochastic dynamical systems. More specifically, we derive a stochastic version of contraction analysis in the specialized context of state-independent metrics.

We prove in section 2 that the mean square distance between any two trajectories of a stochastically contracting system is upper-bounded by a constant after exponential transients. In contrast with previous works on incremental stochastic stability [5], we consider the case when the two trajectories are subject to *distinct* and independent noises, as detailed in section 2.2.1. This specificity enables our theory to have a number of new and practically important applications. However, the fact that the noise does not vanish as two trajectories get very close to each other will prevent us from obtaining asymptotic almost-sure stability results (see section 2.3.2).

In section 3, we show that results on combinations of deterministic contracting systems have simple analogues in the stochastic case. These combination properties allow one to build by recursion stochastically contracting systems of arbitrary size.

Finally, as illustrations of our results, we study in section 4 several examples, including contracting observers with noisy measurements, stochastic composite variables and synchronization phenomena in networks of noisy dynamical systems.

2 Main results

2.1 Background

2.1.1 Nonlinear contraction theory

Contraction theory [24] provides a set of tools to analyze the incremental exponential stability of nonlinear systems, and has been applied notably to observer design [24, 25, 1, 21, 36], synchronization analysis [35, 28] and systems neuroscience modelling [15]. Nonlinear contracting systems enjoy desirable aggregation properties, in that contraction is preserved under many

types of system combinations given suitable simple conditions [24].

While we shall derive global properties of nonlinear systems, many of our results can be expressed in terms of eigenvalues of symmetric matrices [19]. Given a square matrix $\mathbf{A}$, the symmetric part of $\mathbf{A}$ is denoted by $\mathbf{A}_s$. The smallest and largest eigenvalues of $\mathbf{A}_s$ are denoted by $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$. Given these notations, the matrix $\mathbf{A}$ is *positive definite* (denoted $\mathbf{A} > \mathbf{0}$) if $\lambda_{\min}(\mathbf{A}) > 0$, and it is *uniformly* positive definite if

$$\exists \beta > 0 \quad \forall \mathbf{x}, t \quad \lambda_{\min}(\mathbf{A}(\mathbf{x}, t)) \geq \beta$$

The basic theorem of contraction analysis, derived in [24], can be stated as follows

Theorem 1 (Contraction) *Consider, in $\mathbb{R}^n$, the deterministic system*

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) \tag{2.1}$$

where $\mathbf{f}$ is a smooth nonlinear function. Denote the Jacobian matrix of $\mathbf{f}$ with respect to its first variable by $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$. If there exists a square matrix $\mathbf{\Theta}(\mathbf{x}, t)$ such that $\mathbf{M}(\mathbf{x}, t) = \mathbf{\Theta}(\mathbf{x}, t)^T \mathbf{\Theta}(\mathbf{x}, t)$ is uniformly positive definite and the matrix

$$\mathbf{F}(\mathbf{x}, t) = \left(\frac{d}{dt} \mathbf{\Theta}(\mathbf{x}, t) + \mathbf{\Theta}(\mathbf{x}, t) \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right) \mathbf{\Theta}^{-1}(\mathbf{x}, t)$$

is uniformly negative definite, then all system trajectories converge exponentially to a single trajectory, with convergence rate $|\sup_{\mathbf{x},t} \lambda_{\max}(\mathbf{F})| = \lambda > 0$. The system is said to be contracting*, $\mathbf{F}$ is called its* generalized Jacobian*, $\mathbf{M}(\mathbf{x}, t)$ its contraction* metric *and λ its contraction* rate*.*

2.1.2 Standard stochastic stability

In this section, we present very informally the basic ideas of standard stochastic stability (for a rigourous treatment, the reader is referred to e.g. [22]). This will set the context to understand the forthcoming difficulties and differences associated with incremental stochastic stability.

For simplicity, we consider the special case of global exponential stability. Let $\mathbf{x}(t)$ be a Markov stochastic process and assume that there exists a non-negative function V ($V(\mathbf{x})$ may represent e.g. the squared distance of $\mathbf{x}$ from the origin) such that

$$\forall \mathbf{x} \in \mathbb{R}^n \quad \widetilde{A} V(\mathbf{x}) \leq -\lambda V(\mathbf{x}) \tag{2.2}$$

where λ is a positive real number and $\widetilde{A}$ is the infinitesimal operator of the process $\mathbf{x}(t)$. The operator $\widetilde{A}$ is the stochastic analogue of the deterministic differentiation operator. In the case that $\mathbf{x}(t)$ is an Itô process, $\widetilde{A}$ corresponds to the widely-used [27, 34, 8] differential generator $\mathscr{L}$ (for a proof of this fact, see [22], p. 15 or [3], p. 42).

For $\mathbf{x}_0 \in \mathbb{R}^n$, let $\mathbb{E}_{\mathbf{x}_0}(\cdot) = \mathbb{E}(\cdot|\mathbf{x}(0) = \mathbf{x}_0)$. Then by Dynkin's formula ([22], p. 10), one has

$$\begin{array}{rcl} \forall t \geq 0 \quad \mathbb{E}_{\mathbf{x}_0} V(\mathbf{x}(t)) - V(\mathbf{x}_0) & = & \mathbb{E}_{\mathbf{x}_0} \int_0^t \widetilde{A} V(\mathbf{x}(s)) ds \\ & \leq & -\lambda \mathbb{E}_{\mathbf{x}_0} \int_0^t V(\mathbf{x}(s)) ds = -\lambda \int_0^t \mathbb{E}_{\mathbf{x}_0} V(\mathbf{x}(s)) ds \end{array}$$

Applying the Gronwall's lemma to the deterministic real-valued function $t \to \mathbb{E}_{\mathbf{x}_0} V(\mathbf{x}(t))$ yields

$$\forall t \geq 0 \quad \mathbb{E}_{\mathbf{x}_0} V(\mathbf{x}(t)) \leq V(\mathbf{x}_0) e^{-\lambda t}$$

If we assume furthermore that $\mathbb{E}_{\mathbf{x}_0} V(\mathbf{x}(t)) < \infty$ for all t, then the above implies that $V(\mathbf{x}(t))$ is a supermartingale (see lemma 3 in the Appendix for details), which yields, by the supermartingale inequality

$$\mathbb{P}_{\mathbf{x}_0} \left(\sup_{T \leq t < \infty} V(\mathbf{x}(t)) \geq A \right) \leq \frac{\mathbb{E}_{\mathbf{x}_0} V(\mathbf{x}(T))}{A} \leq \frac{V(\mathbf{x}_0) e^{-\lambda T}}{A} \tag{2.3}$$

Thus, one obtains an *almost-sure* stability result, in the sense that

$$\forall A > 0 \quad \lim_{T \to \infty} \mathbb{P}_{\mathbf{x}_0} \left(\sup_{T \leq t < \infty} V(\mathbf{x}(t)) \geq A \right) = 0 \tag{2.4}$$

2.2 The stochastic contraction theorem

2.2.1 Settings

Consider a noisy system described by an Itô stochastic differential equation

$$d\mathbf{a} = \mathbf{f}(\mathbf{a}, t) dt + \sigma(\mathbf{a}, t) dW^d \tag{2.5}$$

where $\mathbf{f}$ is a $\mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}^n$ function, σ is a $\mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}^{nd}$ matrix-valued function and W^d is a standard d-dimensional Wiener process.

To ensure existence and uniqueness of solutions to equation (2.5), we assume, here and in the remainder of the paper, the following standard conditions on $\mathbf{f}$ and σ

Lipschitz condition: There exists a constant $K_1 > 0$ such that

$$\forall t \geq 0, \ \mathbf{a}, \mathbf{b} \in \mathbb{R}^n \quad \|\mathbf{f}(\mathbf{a}, t) - \mathbf{f}(\mathbf{b}, t)| + \|\sigma(\mathbf{a}, t) - \sigma(\mathbf{b}, t)\| \leq K_1 \|\mathbf{a} - \mathbf{b}\|$$

Restriction on growth: There exists a constant $K_2 > 0$

$$\forall t \geq 0, \ \mathbf{a} \in \mathbb{R}^n \quad \|\mathbf{f}(\mathbf{a}, t)\|^2 + \|\sigma(\mathbf{a}, t)\|^2 \leq K_2 (1 + \|\mathbf{a}\|^2)$$

Under these conditions, one can show ([3], p. 105) that equation (2.5) has on $[0, \infty[$ a unique $\mathbb{R}^n$-valued solution $\mathbf{a}(t)$, continuous with probability one, and satisfying the initial condition $\mathbf{a}(0) = \mathbf{a}_0$, with $\mathbf{a}_0 \in \mathbb{R}^n$.

In order to investigate the incremental stability properties of system (2.5), consider now two system trajectories $\mathbf{a}(t)$ and $\mathbf{b}(t)$. Our goal will consist of studying the trajectories $\mathbf{a}(t)$ and $\mathbf{b}(t)$ with respect to each other. For this, we consider the *augmented* system $\mathbf{x}(t) = (\mathbf{a}(t), \mathbf{b}(t))^T$, which follows the equation

$$\begin{aligned} d\mathbf{x} &= \begin{pmatrix} \mathbf{f}(\mathbf{a}, t) \\ \mathbf{f}(\mathbf{b}, t) \end{pmatrix} dt + \begin{pmatrix} \sigma(\mathbf{a}, t) & 0 \\ 0 & \sigma(\mathbf{b}, t) \end{pmatrix} \begin{pmatrix} dW_1^d \\ dW_2^d \end{pmatrix} \\ &= \widehat{\mathbf{f}}(\mathbf{x}, t)dt + \widehat{\sigma}(\mathbf{x}, t)dW^{2d} \end{aligned} \tag{2.6}$$

Important remark As stated in the introduction, the systems $\mathbf{a}$ and $\mathbf{b}$ are driven by *distinct* and independent Wiener processes W_1^d and W_2^d. This makes our approach considerably different from [5], where the authors studied two trajectories driven by *the same* Wiener process.

Our approach enables us to study the stability of the system with respect to variations in initial conditions *and* to random perturbations: indeed, two trajectories of any real-life system are typically affected by distinct “realizations” of the noise. In addition, it leads very naturally to nice results on the comparison of noisy and noise-free trajectories (cf. section 2.4), which are particularly useful in applications (cf. section 4).

However, because of the very fact that the two trajectories are driven by distinct Wiener processes, we cannot expect the influence of the noise to vanish when the two trajectories get very close to each other. This constrasts with [5], and more generally, with the standard stochastic stability case, where the noise vanishes near the origin (cf. section 2.1.2). The consequences of this will be discussed in detail in section 2.3.2.

2.2.2 The basic stochastic contraction theorem

We introduce two hypotheses

(H1) $\mathbf{f}(\mathbf{a}, t)$ is contracting in the identity metric, with contraction rate λ, (i.e. $\forall \mathbf{a}, t \quad \lambda_{\max}\left(\frac{\partial \mathbf{f}}{\partial \mathbf{a}}\right) \le -\lambda$)

(H2) $\mathrm{tr}\left(\sigma(\mathbf{a}, t)^T \sigma(\mathbf{a}, t)\right)$ is uniformly upper-bounded by a constant C (i.e. $\forall \mathbf{a}, t \quad \mathrm{tr}\left(\sigma(\mathbf{a}, t)^T \sigma(\mathbf{a}, t)\right) \le C$)

In other words, **(H1)** says that the noise-free system is *contracting*, while **(H2)** says that the variance of the noise is upper-bounded by a constant.

Definition 1 *A system that verifies* **(H1)** *and* **(H2)** *is said to be* stochastically contracting *in the identity metric, with rate λ and bound C.*

Consider now the Lyapunov-like function $V(\mathbf{x}) = \|\mathbf{a}-\mathbf{b}\|^2 = (\mathbf{a}-\mathbf{b})^T(\mathbf{a}-\mathbf{b})$. Using **(H1)** and **(H2)**, we derive below an inequality on $\widetilde{A}V(\mathbf{x})$, similar to equation (2.2) in section 2.1.2.

Lemma 1 *Under* **(H1)** *and* **(H2)**, *one has the inequality*

$$\widetilde{A}V(\mathbf{x}) \leq -2\lambda V(\mathbf{x}) + 2C \tag{2.7}$$

Proof Since $\mathbf{x}(t)$ is an Itô process, $\widetilde{A}$ is given by the differential operator $\mathscr{L}$ of the process [22, 3]. Thus, by the Itô formula

$$\begin{aligned}
\widetilde{A}V(\mathbf{x}) = \mathscr{L}V(\mathbf{x}) &= \frac{\partial V(\mathbf{x})}{\partial \mathbf{x}}\widehat{\mathbf{f}}(\mathbf{x},t) + \frac{1}{2}\mathrm{tr}\left(\widehat{\sigma}(\mathbf{x},t)^T \frac{\partial^2 V(\mathbf{x})}{\partial \mathbf{x}^2}\widehat{\sigma}(\mathbf{x},t)\right) \\
&= \sum_{1\leq i\leq 2n} \frac{\partial V}{\partial \mathbf{x}_i}\widehat{\mathbf{f}}(\mathbf{x},t)_i + \frac{1}{2}\sum_{1\leq i,j,k\leq 2n} \widehat{\sigma}(\mathbf{x},t)_{ij}\frac{\partial^2 V}{\partial \mathbf{x}_i \partial \mathbf{x}_k}\widehat{\sigma}(\mathbf{x},t)_{kj} \\
&= \sum_{1\leq i\leq n} \frac{\partial V}{\partial \mathbf{a}_i}\mathbf{f}(\mathbf{a},t)_i + \sum_{1\leq i\leq n} \frac{\partial V}{\partial \mathbf{b}_i}\mathbf{f}(\mathbf{b},t)_i \\
&\quad + \frac{1}{2}\sum_{1\leq i,j,k\leq n} \sigma(\mathbf{a},t)_{ij}\frac{\partial^2 V}{\partial \mathbf{a}_i \partial \mathbf{a}_k}\sigma(\mathbf{a},t)_{kj} \\
&\quad + \frac{1}{2}\sum_{1\leq i,j,k\leq n} \sigma(\mathbf{b},t)_{ij}\frac{\partial^2 V}{\partial \mathbf{b}_i \partial \mathbf{b}_k}\sigma(\mathbf{b},t)_{kj} \\
&= 2(\mathbf{a}-\mathbf{b})^T(\mathbf{f}(\mathbf{a},t) - \mathbf{f}(\mathbf{b},t)) \\
&\quad + \mathrm{tr}(\sigma(\mathbf{a},t)^T\sigma(\mathbf{a},t)) + \mathrm{tr}(\sigma(\mathbf{b},t)^T\sigma(\mathbf{b},t))
\end{aligned}$$

Fix $t \geq 0$ and, as in [10], consider the real-valued function

$$r(\mu) = (\mathbf{a}-\mathbf{b})^T(\mathbf{f}(\mu\mathbf{a} + (1-\mu)\mathbf{b},t) - \mathbf{f}(\mathbf{b},t))$$

Since $\mathbf{f}$ is C^1, r is C^1 over $[0,1]$. By the mean value theorem, there exists $\mu_0 \in]0,1[$ such that

$$r'(\mu_0) = r(1) - r(0) = (\mathbf{a}-\mathbf{b})^T(\mathbf{f}(a) - \mathbf{f}(b))$$

On the other hand, one obtains by differentiating r

$$r'(\mu_0) = (\mathbf{a}-\mathbf{b})^T\left(\frac{\partial \mathbf{f}}{\partial \mathbf{a}}(\mu_0\mathbf{a} + (1-\mu_0)\mathbf{b},t)\right)(\mathbf{a}-\mathbf{b})$$

Thus, one has

$$\begin{aligned}
(\mathbf{a}-\mathbf{b})^T(\mathbf{f}(\mathbf{a}) - \mathbf{f}(\mathbf{b})) &= (\mathbf{a}-\mathbf{b})^T\left(\frac{\partial \mathbf{f}}{\partial \mathbf{a}}(\mu_0\mathbf{a} + (1-\mu_0)\mathbf{b},t)\right)(\mathbf{a}-\mathbf{b}) \\
&\leq -\lambda(\mathbf{a}-\mathbf{b})^T(\mathbf{a}-\mathbf{b}) = -2\lambda V(\mathbf{x})
\end{aligned} \tag{2.8}$$

where the inequality is obtained by using **(H1)**.

Finally,

$$\begin{aligned}\widetilde{A}V(\mathbf{x}) &= 2(\mathbf{a}-\mathbf{b})^T(\mathbf{f}(\mathbf{a})-\mathbf{f}(\mathbf{b})) + \mathrm{tr}(\sigma(\mathbf{a},t)^T\sigma(\mathbf{a},t)) + \mathrm{tr}(\sigma(\mathbf{b},t)^T\sigma(\mathbf{b},t)) \\ &\leq -2\lambda V(\mathbf{x}) + 2C\end{aligned}$$

where the inequality is obtained by using **(H2)**. □

We are now in a position to prove our main theorem on stochastic incremental stability.

Theorem 2 (Stochastic contraction) *Assume that system (2.5) verifies* **(H1)** *and* **(H2)***. Let* $\mathbf{a}(t)$ *and* $\mathbf{b}(t)$ *be two trajectories whose initial conditions are given by a probability distribution* $p(\mathbf{x}(0)) = p(\mathbf{a}(0), \mathbf{b}(0))$*. Then*

$$\forall t \geq 0 \quad \mathbb{E}\left(\|\mathbf{a}(t)-\mathbf{b}(t)\|^2\right) \leq \frac{C}{\lambda} + e^{-2\lambda t}\int\left[\|\mathbf{a}_0-\mathbf{b}_0\|^2 - \frac{C}{\lambda}\right]^+ dp(\mathbf{a}_0,\mathbf{b}_0) \tag{2.9}$$

where $[\cdot]^+ = \max(0,\cdot)$*. This implies in particular*

$$\forall t \geq 0 \quad \mathbb{E}\left(\|\mathbf{a}(t)-\mathbf{b}(t)\|^2\right) \leq \frac{C}{\lambda} + \mathbb{E}\left(\|\mathbf{a}(0)-\mathbf{b}(0)\|^2\right)e^{-2\lambda t} \tag{2.10}$$

Proof Let $\mathbf{x}_0 = (\mathbf{a}_0, \mathbf{b}_0) \in \mathbb{R}^{2n}$. By Dynkin's formula ([22], p. 10)

$$\mathbb{E}_{\mathbf{x}_0}V(\mathbf{x}(t)) - V(\mathbf{x}_0) = \mathbb{E}_{\mathbf{x}_0}\int_0^t \widetilde{A}V(\mathbf{x}(s))ds$$

Thus one has $\forall u,t \quad 0 \leq u \leq t < \infty$

$$\begin{aligned}\mathbb{E}_{\mathbf{x}_0}V(\mathbf{x}(t)) - \mathbb{E}_{\mathbf{x}_0}V(\mathbf{x}(u)) &= \mathbb{E}_{\mathbf{x}_0}\int_u^t \widetilde{A}V(\mathbf{x}(s))ds \\ &\leq \mathbb{E}_{\mathbf{x}_0}\int_u^t (-2\lambda V(\mathbf{x}(s)) + 2C)ds \qquad (2.11) \\ &= \int_u^t (-2\lambda \mathbb{E}_{\mathbf{x}_0}V(\mathbf{x}(s)) + 2C)ds\end{aligned}$$

where inequality (2.11) is obtained by using lemma 1.

Denote by $g(t)$ the *deterministic* quantity $\mathbb{E}_{\mathbf{x}_0}V(\mathbf{x}(t))$. Clearly, $g(t)$ is a continuous function of t since $\mathbf{x}(t)$ is a continuous process. The function g then satisfies the conditions of the Gronwall-type lemma 4 in the Appendix, and as a consequence

$$\forall t \geq 0 \quad \mathbb{E}_{\mathbf{x}_0}V(\mathbf{x}(t)) \leq \frac{C}{\lambda} + \left[V(\mathbf{x}_0) - \frac{C}{\lambda}\right]^+ e^{-2\lambda T}$$

Integrating the above inequality with respect to $\mathbf{x}_0$ yields the desired result (2.9). Next, inequality (2.10) follows from (2.9) by remarking that

$$\begin{aligned}\int\left[\|\mathbf{a}_0-\mathbf{b}_0\|^2 - \frac{C}{\lambda}\right]^+ dp(\mathbf{a}_0,\mathbf{b}_0) &\leq \int \|\mathbf{a}_0-\mathbf{b}_0\|^2 dp(\mathbf{a}_0,\mathbf{b}_0) \\ &= \mathbb{E}\left(\|\mathbf{a}(0)-\mathbf{b}(0)\|^2\right) \qquad (2.12)\end{aligned}$$

$\square$

Remark Let $\epsilon > 0$ and $T_\epsilon = \frac{1}{2\lambda} \log \left(\sqrt{\frac{\mathbb{E}(\|\mathbf{a}_0 - \mathbf{b}_0\|^2)}{\epsilon}} \right)$. Then inequality (2.10) and Jensen's inequality [30] imply

$$\forall t \geq T_\epsilon \quad \mathbb{E}(\|\mathbf{a}(t) - \mathbf{b}(t)\|) \leq \sqrt{C/\lambda + \epsilon} \tag{2.13}$$

Since $\|\mathbf{a}(t) - \mathbf{b}(t)\|$ is non-negative, (2.13) together with Markov inequality [11] allow one to obtain the following probabilistic bound on the distance between $\mathbf{a}(t)$ and $\mathbf{b}(t)$

$$\forall A > 0 \ \forall t \geq T_\epsilon \quad \mathbb{P}\left(\|\mathbf{a}(t) - \mathbf{b}(t)\| \geq A\right) \leq \frac{\sqrt{C/\lambda + \epsilon}}{A}$$

Note however that this bound is much weaker than the asymptotic almost-sure bound (2.4).

2.2.3 Generalization to time-varying metrics

Theorem 2 can be vastly generalized by considering general time-dependent metrics (the case of state-dependent metrics is not considered in this article and will be the subject of a future work). Specifically, let us replace **(H1)** and **(H2)** by the following hypotheses

(H1') There exists a uniformly positive definite metric $\mathbf{M}(t) = \mathbf{\Theta}(t)^T \mathbf{\Theta}(t)$, with the lower-bound $\beta > 0$ (i.e. $\forall \mathbf{x}, t \ \mathbf{x}^T \mathbf{M}(t) \mathbf{x} \geq \beta \|\mathbf{x}\|^2$) and $\mathbf{f}(\mathbf{a}, t)$ is contracting in that metric, with contraction rate λ, i.e.

$$\lambda_{\max}\left(\left(\frac{d}{dt} \mathbf{\Theta}(t) + \mathbf{\Theta}(t) \frac{\partial \mathbf{f}}{\partial \mathbf{a}} \right) \mathbf{\Theta}^{-1}(t) \right) \leq -\lambda \quad \text{uniformly}$$

or equivalently

$$\mathbf{M}(t) \frac{\partial \mathbf{f}}{\partial \mathbf{a}} + \left(\frac{\partial \mathbf{f}}{\partial \mathbf{a}} \right)^T \mathbf{M}(t) + \frac{d}{dt} \mathbf{M}(t) \leq -2\lambda \mathbf{M}(t) \quad \text{uniformly}$$

(H2') $\mathrm{tr}\left(\sigma(\mathbf{a}, t)^T \mathbf{M}(t) \sigma(\mathbf{a}, t)\right)$ is uniformly upper-bounded by a constant C

Definition 2 *A system that verifies* **(H1')** *and* **(H2')** *is said to be* stochastically contracting *in the metric* $\mathbf{M}(t)$*, with rate* λ *and bound* C*.*

Consider now the generalized Lyapunov-like function $V_1(\mathbf{x}, t) = (\mathbf{a} - \mathbf{b})^T \mathbf{M}(t)(\mathbf{a} - \mathbf{b})$. Lemma 1 can then be generalized as follows.

Lemma 2 *Under* **(H1')** *and* **(H2')***, one has the inequality*

$$\widetilde{A} V_1(\mathbf{x}, t) \leq -2\lambda V_1(\mathbf{x}, t) + 2C \tag{2.14}$$

Proof Let us compute first $\widetilde{A}V_1$

$$\begin{aligned}\widetilde{A}V_1(\mathbf{x},t) &= \frac{\partial V_1}{\partial t} + \frac{\partial V_1}{\partial \mathbf{x}}\widehat{\mathbf{f}}(\mathbf{x},t) + \frac{1}{2}\mathrm{tr}\left(\widehat{\sigma}(\mathbf{x},t)^T\frac{\partial^2 V_1}{\partial \mathbf{x}^2}\widehat{\sigma}(\mathbf{x},t)\right) \\ &= (\mathbf{a}-\mathbf{b})^T\left(\frac{d}{dt}\mathbf{M}(t)\right)(\mathbf{a}-\mathbf{b}) + 2(\mathbf{a}-\mathbf{b})^T\mathbf{M}(t)(\mathbf{f}(\mathbf{a},t)-\mathbf{f}(\mathbf{b},t)) \\ &\quad +\mathrm{tr}(\sigma(\mathbf{a},t)^T\mathbf{M}(t)\sigma(\mathbf{a},t)) + \mathrm{tr}(\sigma(\mathbf{b},t)^T\mathbf{M}(t)\sigma(\mathbf{b},t))\end{aligned}$$

Fix $t>0$ and consider the real-valued function

$$r(\mu) = (\mathbf{a}-\mathbf{b})^T\mathbf{M}(t)(\mathbf{f}(\mu\mathbf{a}+(1-\mu)\mathbf{b},t)-\mathbf{f}(\mathbf{b},t))$$

Since $\mathbf{f}$ is C^1, r is C^1 over $[0,1]$. By the mean value theorem, there exists $\mu_0 \in]0,1[$ such that

$$r'(\mu_0) = r(1)-r(0) = (\mathbf{a}-\mathbf{b})^T\mathbf{M}(t)(\mathbf{f}(\mathbf{a})-\mathbf{f}(\mathbf{b}))$$

On the other hand, one obtains by differentiating r

$$r'(\mu_0) = (\mathbf{a}-\mathbf{b})^T\mathbf{M}(t)\left(\frac{\partial \mathbf{f}}{\partial \mathbf{a}}(\mu_0\mathbf{a}+(1-\mu_0)\mathbf{b},t)\right)(\mathbf{a}-\mathbf{b})$$

Thus, letting $\mathbf{c}=\mu_0\mathbf{a}+(1-\mu_0)\mathbf{b}$, one has

$$\begin{aligned}&(\mathbf{a}-\mathbf{b})^T\left(\tfrac{d}{dt}\mathbf{M}(t)\right)(\mathbf{a}-\mathbf{b}) + 2(\mathbf{a}-\mathbf{b})^T\mathbf{M}(t)(\mathbf{f}(\mathbf{a})-\mathbf{f}(\mathbf{b})) \\ &= (\mathbf{a}-\mathbf{b})^T\left(\tfrac{d}{dt}\mathbf{M}(t)\right)(\mathbf{a}-\mathbf{b}) + 2(\mathbf{a}-\mathbf{b})^T\mathbf{M}(t)\left(\tfrac{\partial \mathbf{f}}{\partial \mathbf{a}}(\mathbf{c},t)\right)(\mathbf{a}-\mathbf{b}) \\ &= (\mathbf{a}-\mathbf{b})^T\left(\tfrac{d}{dt}\mathbf{M}(t) + \mathbf{M}(t)\left(\tfrac{\partial \mathbf{f}}{\partial \mathbf{a}}(\mathbf{c},t)\right) + \left(\tfrac{\partial \mathbf{f}}{\partial \mathbf{a}}(\mathbf{c},t)\right)^T\mathbf{M}(t)\right)(\mathbf{a}-\mathbf{b}) \\ &\leq -2\lambda(\mathbf{a}-\mathbf{b})^T\mathbf{M}(t)(\mathbf{a}-\mathbf{b}) = -2\lambda V_1(\mathbf{x}) \qquad (2.15)\end{aligned}$$

where the inequality is obtained by using **(H1')**.

Finally, combining equation (2.15) with **(H2')** allows to obtain the desired result. □

We can now state the generalized stochastic contraction theorem

Theorem 3 (Generalized stochastic contraction) *Assume that system (2.5) verifies* **(H1')** *and* **(H2')**. *Let* $\mathbf{a}(t)$ *and* $\mathbf{b}(t)$ *be two trajectories whose initial conditions are given by a probability distribution* $p(\mathbf{x}(0)) = p(\mathbf{a}(0),\mathbf{b}(0))$. *Then*

$$\begin{aligned}&\forall t\geq 0 \quad \mathbb{E}\left((\mathbf{a}(t)-\mathbf{b}(t))^T\mathbf{M}(t)(\mathbf{a}(t)-\mathbf{b}(t))\right) \leq \\ &\frac{C}{\lambda} + e^{-2\lambda t}\int\left[(\mathbf{a}_0-\mathbf{b}_0)^T\mathbf{M}(0)(\mathbf{a}_0-\mathbf{b}_0) - \frac{C}{\lambda}\right]^+ dp(\mathbf{a}_0,\mathbf{b}_0) \qquad (2.16)\end{aligned}$$

In particular,

$$\forall t\geq 0 \quad \mathbb{E}\left(\|\mathbf{a}(t)-\mathbf{b}(t)\|^2\right) \leq \frac{1}{\beta}\left(\frac{C}{\lambda} + \mathbb{E}\left(\|\mathbf{a}(0)-\mathbf{b}(0)\|^2\right)e^{-2\lambda t}\right) \qquad (2.17)$$

Proof Following the same reasoning as in the proof of theorem 2, one obtains

$$\forall t \geq 0 \quad \mathbb{E}_{\mathbf{x}_0} V_1(\mathbf{x}(t)) \leq \frac{C}{\lambda} + \left[V_1(\mathbf{x}_0) - \frac{C}{\lambda} \right]^+ e^{-2\lambda t}$$

which leads to (2.16) by integrating with respect to $(\mathbf{a}_0, \mathbf{b}_0)$. Next, observing that

$$\|\mathbf{a}(t) - \mathbf{b}(t)\|^2 \leq \frac{1}{\beta}(\mathbf{a}(t) - \mathbf{b}(t))^T \mathbf{M}(t)(\mathbf{a}(t) - \mathbf{b}(t)) = \frac{1}{\beta}\mathbb{E}V_1(\mathbf{x}(t))$$

and using the same bounding as in (2.12) lead to (2.17). □

2.3 Strength of the stochastic contraction theorem

2.3.1 "Optimality" of the mean square bound

Consider the following linear dynamical system, known as the Ornstein-Uhlenbeck (colored noise) process

$$da = -\lambda a dt + \sigma dW \tag{2.18}$$

Clearly, the noise-free system is contracting with rate λ and the trace of the noise matrix is upper-bounded by σ^2. Let $a(t)$ and $b(t)$ be two system trajectories starting respectively at a_0 and b_0 (deterministic initial conditions). Then by theorem 2, we have

$$\forall t \geq 0 \quad \mathbb{E}\left((a(t) - b(t))^2\right) \leq \frac{\sigma^2}{\lambda} + \left[(a_0 - b_0)^2 - \frac{\sigma^2}{\lambda}\right]^+ e^{-2\lambda t} \tag{2.19}$$

Let us verify this result by solving directly equation (2.18). The solution of equation (2.18) is ([3], p. 134)

$$a(t) = a_0 e^{-\lambda t} + \sigma \int_0^t e^{\lambda(s-t)} dW(s) \tag{2.20}$$

Next, let us compute the mean square distance between the two trajectories $a(t)$ and $b(t)$

$$\begin{aligned}
\mathbb{E}((a(t) - b(t))^2) &= (a_0 - b_0)^2 e^{-2\lambda t} + \\
&\quad \sigma^2 \left(\mathbb{E}\left(\left(\int_0^t e^{\lambda(s-t)} dW_1(s) \right)^2 \right) + \mathbb{E}\left(\left(\int_0^t e^{\lambda(u-t)} dW_2(u) \right)^2 \right) \right) \\
&= (a_0 - b_0)^2 e^{-2\lambda t} + \frac{\sigma^2}{\lambda}(1 - e^{-2\lambda t}) \\
&\leq \frac{\sigma^2}{\lambda} + \left[(a_0 - b_0)^2 - \frac{\sigma^2}{\lambda}\right]^+ e^{-2\lambda t}
\end{aligned}$$

The last inequality is in fact an equality when $(a_0 - b_0)^2 \geq \frac{\sigma^2}{\lambda}$. Thus, this calculation shows that the upper-bound (2.19) given by theorem 2 is optimal, in the sense that it can be attained.

2.3.2 No asymptotic almost-sure stability

From the explicit form (2.20) of the solutions, one can deduce that the distributions of $a(t)$ and $b(t)$ converge to the normal distribution $\mathscr{N}\left(0, \frac{\sigma^2}{2\lambda}\right)$ ([3], p. 135). Since $a(t)$ and $b(t)$ are independent, the distribution of the difference $a(t)-b(t)$ will then converge to $\mathscr{N}\left(0, \frac{\sigma^2}{\lambda}\right)$. This observation shows that, contrary to the case of standard stochastic stability (cf. section 2.1.2), one cannot – in general – obtain asymptotic almost-sure incremental stability results (which would imply that the distribution of the difference converges instead to the constant 0).

Compare indeed equations (2.2) (the condition for standard stability, section 2.1.2) and (2.7) (the condition for incremental stability, section 2.2.2). The difference lies in the term $2C$, which stems from the fact that the influence of the noise does not vanish when two trajectories get very close to each other (cf. section 2.2.1). The presence of this extra term prevents $\widetilde{A}V(\mathbf{x}(t))$ from being always non-positive, and as a result, it prevents $V(\mathbf{x}(t))$ from being always "non-increasing". As a consequence, $V(\mathbf{x}(t))$ is not – in general – a supermartingale, and one cannot then use the supermartingale inequality to obtain asymptotic almost-sure bounds, as in equation (2.3).

Remark If one is interested in *finite time* bounds then the supermartingale inequality is still applicable, see ([22], p. 86) for details.

2.4 Noisy and noise-free trajectories

Consider the following augmented system

$$d\mathbf{x} = \begin{pmatrix} \mathbf{f}(\mathbf{a},t) \\ \mathbf{f}(\mathbf{b},t) \end{pmatrix} dt + \begin{pmatrix} 0 & 0 \\ 0 & \sigma(\mathbf{b},t) \end{pmatrix} \begin{pmatrix} dW_d^1 \\ dW_d^2 \end{pmatrix} = \widehat{\mathbf{f}}(\mathbf{x},t)dt + \widehat{\sigma}(\mathbf{x},t)dW_{2d} \tag{2.21}$$

This equation is the same as equation (2.6) except that the $\mathbf{a}$-system is not perturbed by noise. Thus $V(\mathbf{x}) = \|\mathbf{a}-\mathbf{b}\|^2$ will represent the distance between a noise-free trajectory and a noisy one. All the calculations will be the same as in the previous development, with C being replaced by $C/2$. One can then derive the following corollary

Corollary 1 *Assume that system (2.5) verifies* **(H1')** *and* **(H2')**. *Let* $\mathbf{a}(t)$ *be a* noise-free *trajectory starting at* $\mathbf{a}_0$ *and* $\mathbf{b}(t)$ *a* noisy *trajectory whose initial condition is given by a probability distribution* $p(\mathbf{b}(0))$. *Then*

$$\forall t \geq 0 \quad \mathbb{E}\left(\|\mathbf{a}(t)-\mathbf{b}(t)\|^2\right) \leq \frac{1}{\beta}\left(\frac{C}{2\lambda} + \mathbb{E}\left(\|\mathbf{a}_0-\mathbf{b}(0)\|^2\right)e^{-2\lambda t}\right) \tag{2.22}$$

Remarks

- One can note here that the derivation of corollary 1 is only permitted by our initial choice of considering *distinct* driving Wiener process for the **a**- and **b**-systems (cf. section 2.2.1).
- Corollary 1 provides a *robustness* result for contracting systems, in the sense that any contracting system is *automatically* protected against noise, as quantified by (2.22). This robustness could be related to the exponential nature of contraction stability.

3 Combinations of contracting stochastic systems

Stochastic contraction inherits naturally from deterministic contraction [24] its convenient combination properties. Because contraction is a state-space concept, such properties can be expressed in more general forms than input-output analogues such as passivity-based combinations [29]. The following combination properties allow one to build by recursion stochastically contracting systems of arbitrary size.

Parallel combination Consider two stochastic systems of the same dimension

$$\begin{cases} d\mathbf{x}_1 = \mathbf{f}_1(\mathbf{x}_1, t)dt + \sigma_1(\mathbf{x}_1, t)dW_1 \\ d\mathbf{x}_2 = \mathbf{f}_2(\mathbf{x}_2, t)dt + \sigma_2(\mathbf{x}_2, t)dW_2 \end{cases}$$

Assume that both systems are stochastically contracting in the same *constant* metric $\mathbf{M}$, with rates λ_1 and λ_2 and with bounds C_1 and C_2. Consider a uniformly positive bounded superposition

$$\alpha_1(t)\mathbf{x}_1 + \alpha_2(t)\mathbf{x}_2$$

where $\forall t \geq 0,\ l_i \leq \alpha_i(t) \leq m_i$ for some $l_i, m_i > 0,\ i = 1, 2$.

Clearly, this superposition is stochastically contracting in the metric $\mathbf{M}$, with rate $l_1\lambda_1 + l_2\lambda_2$ and bound $m_1C_1 + m_2C_2$.

Negative feedback combination In this and the following paragraphs, we describe combinations properties for contracting systems in constant metrics $\mathbf{M}$. The case of time-varying metrics can be easily adapted from this development but is skipped here for the sake of clarity.

Consider two coupled stochastic systems

$$\begin{cases} d\mathbf{x}_1 = \mathbf{f}_1(\mathbf{x}_1, \mathbf{x}_2, t)dt + \sigma_1(\mathbf{x}_1, t)dW_1 \\ d\mathbf{x}_2 = \mathbf{f}_2(\mathbf{x}_1, \mathbf{x}_2, t)dt + \sigma_2(\mathbf{x}_2, t)dW_2 \end{cases}$$

Assume that system i ($i = 1, 2$) is stochastically contracting with respect to $\mathbf{M}_i = \mathbf{\Theta}_i^T\mathbf{\Theta}_i$, with rate λ_i and bound C_i.

Assume furthermore that the two systems are connected by *negative feedback* [33]. More precisely, the Jacobian matrices of the couplings are of the form $\mathbf{\Theta}_1\mathbf{J}_{12}\mathbf{\Theta}_2^{-1} = -k\mathbf{\Theta}_2\mathbf{J}_{21}^T\mathbf{\Theta}_1^{-1}$, with k a positive constant. Hence, the Jacobian matrix of the augmented system is given by

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_1 & -k\mathbf{\Theta}_1^{-1}\mathbf{\Theta}_2\mathbf{J}_{21}^T\mathbf{\Theta}_1^{-1}\mathbf{\Theta}_2 \\ \mathbf{J}_{21} & \mathbf{J}_2 \end{pmatrix}$$

Consider a coordinate transform $\mathbf{\Theta} = \begin{pmatrix} \mathbf{\Theta}_1 & \mathbf{0} \\ \mathbf{0} & \sqrt{k}\mathbf{\Theta}_2 \end{pmatrix}$ associated with the metric $\mathbf{M} = \mathbf{\Theta}^T\mathbf{\Theta} > \mathbf{0}$. After some calculations, one has

$$\begin{aligned} \left(\mathbf{\Theta}\mathbf{J}\mathbf{\Theta}^{-1}\right)_s &= \begin{pmatrix} \left(\mathbf{\Theta}_1\mathbf{J}_1\mathbf{\Theta}_1^{-1}\right)_s & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{\Theta}_2\mathbf{J}_2\mathbf{\Theta}_2^{-1}\right)_s \end{pmatrix} \\ &\leq \max(-\lambda_1, -\lambda_2)\mathbf{I} \quad \text{uniformly} \end{aligned} \tag{3.1}$$

The augmented system is thus stochastically contracting in the metric $\mathbf{M}$, with rate $\min(\lambda_1, \lambda_2)$ and bound $C_1 + kC_2$.

Hierarchical combination We first recall a standard result in matrix analysis [19]. Let $\mathbf{A}$ be a symmetric matrix in the form $\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_{21}^T \\ \mathbf{A}_{21} & \mathbf{A}_2 \end{pmatrix}$. Assume that $\mathbf{A}_1$ and $\mathbf{A}_2$ are definite positive. Then $\mathbf{A}$ is definite positive if $\operatorname{sing}^2(\mathbf{A}_{21}) < \lambda_{\min}(\mathbf{A}_1)\lambda_{\min}(\mathbf{A}_2)$ where $\operatorname{sing}(\mathbf{A}_{21})$ denotes the largest singular value of $\mathbf{A}_{21}$. In this case, the smallest eigenvalue of $\mathbf{A}$ satisfies

$$\lambda_{\min}(\mathbf{A}) \geq \frac{\lambda_{\min}(\mathbf{A}_1) + \lambda_{\min}(\mathbf{A}_2)}{2} - \sqrt{\left(\frac{\lambda_{\min}(\mathbf{A}_1) - \lambda_{\min}(\mathbf{A}_2)}{2}\right)^2 + \operatorname{sing}^2(\mathbf{A}_{21})}$$

Consider now the same set-up as in the previous paragraph, except that the connection is now *hierarchical* and upper-bounded. More precisely, the Jacobians of the couplings verify $\mathbf{J}_{12} = \mathbf{0}$ and $\operatorname{sing}^2(\mathbf{\Theta}_2\mathbf{J}_{21}\mathbf{\Theta}_1^{-1}) \leq K$. The Jacobian matrix of the augmented system is then given by $\mathbf{J} = \begin{pmatrix} \mathbf{J}_1 & \mathbf{0} \\ \mathbf{J}_{21} & \mathbf{J}_2 \end{pmatrix}$. Consider a coordinate transform $\mathbf{\Theta}_\epsilon = \begin{pmatrix} \mathbf{\Theta}_1 & \mathbf{0} \\ \mathbf{0} & \epsilon\mathbf{\Theta}_2 \end{pmatrix}$ associated with the metric $\mathbf{M}_\epsilon = \mathbf{\Theta}_\epsilon^T\mathbf{\Theta}_\epsilon > \mathbf{0}$. After some calculations, one has

$$\left(\mathbf{\Theta}\mathbf{J}\mathbf{\Theta}^{-1}\right)_s = \begin{pmatrix} \left(\mathbf{\Theta}_1\mathbf{J}_1\mathbf{\Theta}_1^{-1}\right)_s & \frac{1}{2}\epsilon(\mathbf{\Theta}_2\mathbf{J}_{21}\mathbf{\Theta}_1^{-1})^T \\ \frac{1}{2}\epsilon\mathbf{\Theta}_2\mathbf{J}_{21}\mathbf{\Theta}_1^{-1} & \left(\mathbf{\Theta}_2\mathbf{J}_2\mathbf{\Theta}_2^{-1}\right)_s \end{pmatrix}$$

Set now $\epsilon = \sqrt{\frac{2\lambda_1\lambda_2}{K}}$. The augmented system is then stochastically contracting in the metric $\mathbf{M}_\epsilon$, with rate $\frac{1}{2}(\lambda_1 + \lambda_2 - \sqrt{\lambda_1^2 + \lambda_2^2}))$ and bound $C_1 + \frac{2C_2\lambda_1\lambda_2}{K}$.

Small gains In this paragraph, we require no specific assumption on the form of the couplings. Consider the coordinate transform $\mathbf{\Theta} = \begin{pmatrix} \mathbf{\Theta}_1 & \mathbf{0} \\ \mathbf{0} & \sqrt{k}\mathbf{\Theta}_2 \end{pmatrix}$ associated with the metric $\mathbf{M}_k = \mathbf{\Theta}_k^T \mathbf{\Theta}_k > \mathbf{0}$. Aftersome calculations, one has

$$\left(\mathbf{\Theta}_k \mathbf{J} \mathbf{\Theta}_k^{-1}\right)_s \quad = \quad \begin{pmatrix} \left(\mathbf{\Theta}_1 \mathbf{J}_1 \mathbf{\Theta}_1^{-1}\right)_s & \mathbf{B}_k^T \\ \mathbf{B}_k & \left(\mathbf{\Theta}_2 \mathbf{J}_2 \mathbf{\Theta}_2^{-1}\right)_s \end{pmatrix}$$

where $\mathbf{B}_k = \frac{1}{2}\left(\sqrt{k}\mathbf{\Theta}_2 \mathbf{J}_{21} \mathbf{\Theta}_1^{-1} + \frac{1}{\sqrt{k}}\left(\mathbf{\Theta}_1 \mathbf{J}_{12} \mathbf{\Theta}_2^{-1}\right)^T\right)$.

Following the matrix analysis result stated at the beginning of the previous paragraph, if $\inf_{k>0} \operatorname{sing}^2(\mathbf{B}_k) < \lambda_1 \lambda_2$ then the augmented system is stochastically contracting in the metric $\mathbf{M}_k$, with bound $C_1 + kC_2$ and rate λ verifying

$$\lambda \geq \frac{\lambda_1 + \lambda_2}{2} - \sqrt{\left(\frac{\lambda_1 - \lambda_2}{2}\right)^2 + \inf_{k>0} \operatorname{sing}^2(\mathbf{B}_k)} \tag{3.2}$$

4 Some examples

4.1 Effect of measurement noise on contracting observers

Consider a nonlinear dynamical system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) \tag{4.1}$$

If a measurement $\mathbf{y} = \mathbf{y}(\mathbf{x})$ is available, then it may be possible to choose an output injection matrix $\mathbf{K}(t)$ such that the dynamics

$$\dot{\hat{\mathbf{x}}} = \mathbf{f}(\hat{\mathbf{x}}, t) + \mathbf{K}(t)(\hat{\mathbf{y}} - \mathbf{y}) \tag{4.2}$$

is contracting, with $\hat{\mathbf{y}} = \mathbf{y}(\hat{\mathbf{x}})$. Since the actual state $\mathbf{x}$ is a particular solution of (4.2), any solution $\hat{\mathbf{x}}$ of (4.2) will then converge towards $\mathbf{x}$ exponentially.

Assume now that the measurements are corrupted by additive "white noise". In the case of *linear* measurement, the measurement equation becomes $\mathbf{y} = \mathbf{H}(t)\mathbf{x} + \Sigma(t)\xi(t)$ where $\xi(t)$ is a multidimensional "white noise" and $\Sigma(t)$ is the matrix of measurement noise intensities.

The observer equation is now given by the following Itô stochastic differential equation (using the formal rule $dW = \xi dt$)

$$d\hat{\mathbf{x}} = (\mathbf{f}(\hat{\mathbf{x}}, t) + \mathbf{K}(t)(\mathbf{H}(t)\mathbf{x} - \mathbf{H}(t)\hat{\mathbf{x}}))dt + \mathbf{K}(t)\Sigma(t)dW \tag{4.3}$$

Next, remark that the solution $\mathbf{x}$ of system (4.1) is a also a solution of the noise-free version of system (4.3). By corollary 1, one then has, for any solution $\hat{\mathbf{x}}$ of system (4.3)

$$\forall t \geq 0 \quad \mathbb{E}\left(\|\hat{\mathbf{x}}(t) - \mathbf{x}(t)\|^2\right) \leq \frac{C}{2\lambda} + \|\hat{\mathbf{x}}_0 - \mathbf{x}_0\|^2 e^{-2\lambda t} \tag{4.4}$$

where

$$\lambda = \inf_{\mathbf{x},t} \left| \lambda_{\max}\left(\frac{\partial \mathbf{f}(\mathbf{x},t)}{\partial \mathbf{x}} - \mathbf{K}(t)\mathbf{H}(t)\right)\right|$$
$$C = \sup_{t\geq 0} \operatorname{tr}\left(\Sigma(t)^T \mathbf{K}(t)^T \mathbf{K}(t)\Sigma(t)\right)$$

Remark The choice of the injection gain $\mathbf{K}(t)$ is governed by a trade-off between convergence speed (λ) and noise sensitivity (C/λ) as quantified by (4.4). More generally, the explicit computation of the bound on the expected quadratic estimation error given by (4.4) may open the possibility of *measurement selection* in a way similar to the linear case. If several possible measurements or sets of measurements can be performed, one may try at each instant (or at each step, in a discrete version) to select the most relevant, i.e., the measurement or set of measurements which will best contribute to improving the state estimate. Similarly to the Kalman filters used in [9] for linear systems, this can be achieved by computing, along with the state estimate itself, the corresponding bounds on the expected quadratic estimation error, and then selecting accordingly the measurement which will minimize it.

4.2 Estimation of velocity using composite variables

In this section, we present a very simple example that hopefully suggests the many possibilities that could stem from the combination of our stochastic stability analysis with the composite variables framework [31].

Let x be the position of a mobile subject to a sinusoidal forcing

$$\ddot{x} = -U_1\omega^2 \sin(\omega t) + 2U_2$$

where U_1 and ω are *known* parameters. We would like to compute good approximations of the mobile's velocity v and acceleration a using only measurements of x and without using any filter. For this, construct the following observer

$$\begin{aligned} \frac{d}{dt}\begin{pmatrix} \overline{v} \\ \overline{a} \end{pmatrix} &= \begin{pmatrix} -\alpha_v & 1 \\ -\alpha_a & 0 \end{pmatrix}\begin{pmatrix} \overline{v} \\ \overline{a} \end{pmatrix} + \begin{pmatrix} (\alpha_a - \alpha_v^2)x \\ -\alpha_a\alpha_v x - U_1\omega^3\cos(\omega t) \end{pmatrix} \\ &= \mathbf{A}\begin{pmatrix} \overline{v} \\ \overline{a} \end{pmatrix} + \mathbf{B}\begin{pmatrix} x \\ x \end{pmatrix} - \begin{pmatrix} 0 \\ U_1\omega^3\cos(\omega t) \end{pmatrix} \end{aligned} \tag{4.5}$$

and introduce the composite variables $\widehat{v} = \overline{v} + \alpha_v x$ and $\widehat{a} = \overline{a} + \alpha_a x$. By construction, these variables follow the equation

$$\frac{d}{dt}\begin{pmatrix} \widehat{v} \\ \widehat{a} \end{pmatrix} = \mathbf{A}\begin{pmatrix} \widehat{v} - v \\ \widehat{a} \end{pmatrix} + \begin{pmatrix} 0 \\ -U_1\omega^3\cos(\omega t) \end{pmatrix} \tag{4.6}$$

and therefore, a particular solution of $(\widehat{v}, \widehat{a})$ is clearly (v, a). Choose now $\alpha_a = \alpha_v^2 = \alpha^2$ and let $\mathbf{M}_\alpha = \frac{1}{2}\begin{pmatrix} \alpha^2 & -\alpha/2 \\ -\alpha/2 & 1 \end{pmatrix}$. One can then show that system (4.6) is contracting with rate $\lambda_\alpha = \alpha/2$ in the metric $\mathbf{M}_\alpha$. Thus, by the basic contraction theorem [24], $(\widehat{v}, \widehat{a})$ converges exponentially to (v, a) with rate λ_α in the metric $\mathbf{M}_\alpha$. Also note that the β-bound corresponding to the metric $\mathbf{M}_\alpha$ is given by $\beta_\alpha = \frac{1+\alpha^2-\sqrt{\alpha^4-\alpha^2+1}}{4}$.

Next, assume that the measurements of x are corrupted by additive "white noise", so that $x_{\text{measured}} = x + \sigma\xi$. Equation (4.5) then becomes an Itô stochastic differential equation

$$d\begin{pmatrix} \overline{v} \\ \overline{a} \end{pmatrix} = \left[\mathbf{A}\begin{pmatrix} \overline{v} \\ \overline{a} \end{pmatrix} + \mathbf{B}\begin{pmatrix} x \\ x \end{pmatrix} - \begin{pmatrix} 0 \\ U_1\omega^3\cos(\omega t) \end{pmatrix}\right] dt + \mathbf{B}\begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix} dW$$

By definition of $\mathbf{B}$, the variance of the noise in the metric $\mathbf{M}_\alpha$ is upper-bounded by $\frac{\alpha^6\sigma^2}{2}$. Thus, using again corollary 1, one obtains (see Figure 1 for a numerical simulation)

$$\forall t \geq 0 \quad \mathbb{E}\left(\|\widehat{v}(t) - v(t)\|^2 + \|\widehat{a}(t) - a(t)\|^2\right) \leq \frac{\alpha^5\sigma^2}{2\beta_\alpha} + \frac{\|\widehat{v}_0 - v_0\|^2 + \|\widehat{a}_0 - a_0\|^2}{2\beta_\alpha} e^{-\alpha t}$$

4.3 Stochastic synchronization

Consider a network of n dynamical elements coupled through diffusive connections

$$d\mathbf{x}_i = \left(\mathbf{f}(\mathbf{x}_i, t) + \sum_{j \neq i} \mathbf{K}_{ij}(\mathbf{x}_j - \mathbf{x}_i)\right) dt + \sigma_i(\mathbf{x}_i, t) dW_i^d \qquad i = 1, \ldots, n \tag{4.7}$$

Let

$$\widehat{\mathbf{x}} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}, \quad \widehat{\mathbf{f}}(\widehat{\mathbf{x}}, t) = \begin{pmatrix} \mathbf{f}(\mathbf{x}_1, t) \\ \vdots \\ \mathbf{f}(\mathbf{x}_n, t) \end{pmatrix}, \quad \widehat{\sigma}(\widehat{\mathbf{x}}, t) = \begin{pmatrix} \sigma_1(\mathbf{x}_1, t) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_n(\mathbf{x}_n, t) \end{pmatrix}$$

The global state $\widehat{\mathbf{x}}$ then follows the equation

$$d\widehat{\mathbf{x}} = \left(\widehat{\mathbf{f}}(\widehat{\mathbf{x}}, t) - \mathbf{L}\widehat{\mathbf{x}}\right) dt + \widehat{\sigma}(\widehat{\mathbf{x}}, t) dW^{nd} \tag{4.8}$$

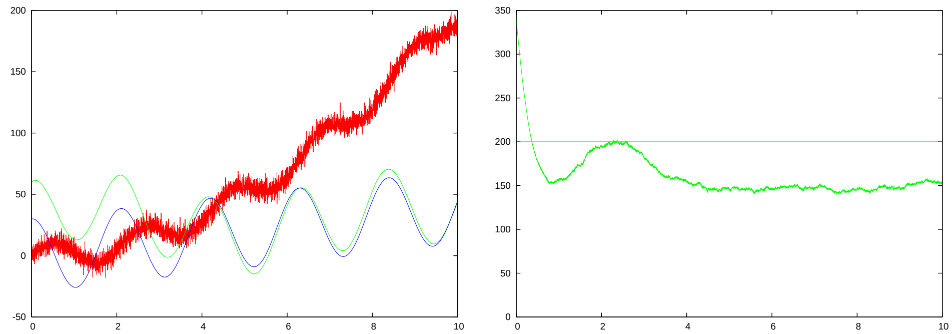

Figure 1: Estimation of the velocity of a mobile using noisy measurements of its position. The simulation was performed using the Euler-Maruyama algorithm [18] with the following parameters: $U_1 = 10$, $U_2 = 2$, $\omega = 3$, $\sigma = 10$ and $\alpha = 1$. Left plot: simulation for one trial. The plot shows the measured position (red), the actual velocity (blue) and the estimate of the velocity using the measured position (green). Right plot: the average over 1000 trials of the squared error $\|\widehat{v} - v\|^2 + \|\widehat{a} - a\|^2$ (green) and the asymptotic bound $\left(\frac{\alpha^5\sigma^2}{2\beta_\alpha} = 200\right)$ given by our approach (red).

In the sequel, we follow the reasoning of [28], which starts by defining an appropriate orthonormal matrix $\mathbf{V}$ describing the synchronization subspace ($\mathbf{V}$ represents the state projection on the subspace $\mathcal{M}^\perp$, orthogonal to the synchronization subspace $\mathcal{M} = \{(\mathbf{x}_1, \ldots, \mathbf{x}_n)^T : \mathbf{x}_1 = \ldots = \mathbf{x}_n\}$, see [28] for details). Denote by $\widehat{\mathbf{y}}$ the state of the projected system, $\widehat{\mathbf{y}} = \mathbf{V}\widehat{\mathbf{x}}$. Since the mapping is linear, Itô differentiation rule simply yields

$$\begin{aligned} d\widehat{\mathbf{y}} &= \mathbf{V}d\widehat{\mathbf{x}} = \left(\mathbf{V}\widehat{\mathbf{f}}(\widehat{\mathbf{x}}, t) - \mathbf{V}\mathbf{L}\widehat{\mathbf{x}}\right) dt + \mathbf{V}\widehat{\sigma}(\widehat{\mathbf{x}}, t) dW^{nd} \\ &= \left(\mathbf{V}\widehat{\mathbf{f}}(\mathbf{V}^T\widehat{\mathbf{y}}, t) - \mathbf{V}\mathbf{L}\mathbf{V}^T\widehat{\mathbf{y}}\right) dt + \mathbf{V}\widehat{\sigma}(\mathbf{V}^T\widehat{\mathbf{y}}, t) dW^{nd} \end{aligned} \tag{4.9}$$

Assume now that $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$ is uniformly upper-bounded. Then for strong enough coupling strength, $\mathbf{A} = \mathbf{V}\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\mathbf{V}^T - \mathbf{V}\mathbf{L}\mathbf{V}^T$ will be uniformly negative definite. Let $\lambda = |\lambda_{\max}(\mathbf{A})| > 0$. System (4.9) then verifies condition **(H1)** with rate λ. Assume furthermore that each noise intensity σ_i is upper-bounded by a constant C_i (i.e. $\sup_{\mathbf{x},t} \mathrm{tr}(\sigma_i(\mathbf{x}, t)^T \sigma_i(\mathbf{x}, t)) \leq C_i$). Condition **(H2)** will then be satisfied with the bound $C = \sum_i C_i$.

Next, consider a noise-free trajectory $\widehat{\mathbf{y}}_u(t)$ of system (4.9). By theorem 3 of [28], we know that $\widehat{\mathbf{y}}_u(t)$ converges exponentially to zero. Thus, by corollary 1, one can conclude that, after exponential transients of rate λ, $\mathbb{E}(\|\widehat{\mathbf{y}}(t)\|^2) \leq \frac{C}{2\lambda}$.

On the other hand, one can show that

$$\|\widehat{\mathbf{y}}(t)\|^2 = \frac{1}{n} \sum_{i,j} \|\mathbf{x}_i - \mathbf{x}_j\|^2$$

Thus, after exponential transients of rate λ, we have

$$\sum_{i,j} \|\mathbf{x}_i - \mathbf{x}_j\|^2 \leq \frac{nC}{2\lambda}$$

Remarks

- The above development is fully compatible with the concurrent synchronization framework [28]. It can also be easily generalized to the case of time-varying metrics by combining theorem 3 of this paper and corollary 1 of [28].
- The synchronization of Itô dynamical systems has been investigated in [4]. However, the systems considered by the authors of that article were *dissipative*. Here, we make a less restrictive assumption, namely, we only require $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$ to be uniformly upper-bounded. This enables us to study the synchronization of a broader class of dynamical systems, which can include nonlinear oscillators or even chaotic systems.

Example As illustration of the above development, we provide here a detailed analysis for the synchronization of noisy FitzHugh-Nagumo oscillators (see [35] for the references). The dynamics of two diffusively-coupled noisy FitzHugh-Nagumo oscillators can be described by

$$\left\{ \begin{array}{l} dv_i = (c(v_i + w_i - \frac{1}{3}v_i^3 + I_i) + k(v_0 - v_i))dt + \sigma dW_i \\ dw_i = -\frac{1}{c}(v_i - a + bw_i)dt \end{array} \right. \qquad i = 1, 2$$

Let $\mathbf{x} = (v_1, w_1, v_2, w_2)^T$ and $\mathbf{V} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}$. The Jacobian matrix of the projected noise-free system is then given by

$$\begin{pmatrix} c - \frac{c(v_1^2 + v_2^2)}{2} - k & c \\ -1/c & -b/c \end{pmatrix}$$

Thus, if the coupling strength verifies $k > c$ then the projected system will be stochastically contracting in the diagonal metric $\mathbf{M} = \mathrm{diag}(1, c)$ with rate $\min(k - c, b/c)$ and bound σ^2. Hence, the average absolute difference between the two "membrane potentials" $|v_1 - v_2|$ will be upper-bounded by $\sigma / \sqrt{\min(1, c) \min(k - c, b/c)}$ (see Figure 2 for a numerical simulation).

Acknowledgments We are grateful to Dr S. Darses, Prof D. Bennequin and Prof M. Yor for stimulating discussions, and to the Associate Editor and the reviewers for their helpful comments.

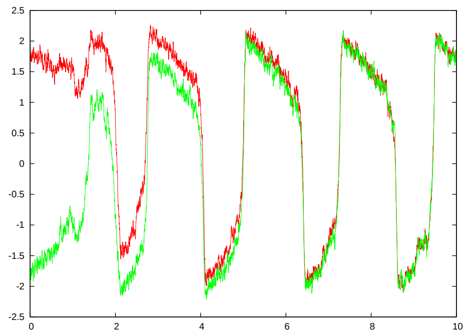

Figure 2: Synchronization of two noisy FitzHugh-Nagumo oscillators. The simulation was performed using the Euler-Maruyama algorithm [18] with the following parameters: $a = 0.3$, $b = 0.2$, $c = 30$, $k = 40$ and $\sigma = 1$. The plot shows the "membrane potentials" of the two oscillators.

A Appendix

A.1 Proof of the supermartingale property

Lemma 3 *Consider a Markov stochastic process* $\mathbf{x}(t)$ *and a non-negative function* V *such that* $\forall t \geq 0 \quad \mathbb{E}V(\mathbf{x}(t)) < \infty$ *and*

$$\forall \mathbf{x} \in \mathbb{R}^n \quad \widetilde{A}V(\mathbf{x}) \leq -\lambda V(\mathbf{x}) \tag{A.1}$$

where λ *is a non-negative real number and* $\widetilde{A}$ *is the infinitesimal operator of the process* $\mathbf{x}(t)$*. Then* $V(\mathbf{x}(t))$ *is a supermartingale with respect to the canonical filtration* $\mathscr{F}_t = \{\mathbf{x}(s), s \leq t\}$*.*

We need to show that for all $s \geq t$, one has $\mathbb{E}(V(\mathbf{x}(s))|\mathscr{F}_t) \leq V(\mathbf{x}(t))$. Since $\mathbf{x}(t)$ is a Markov process, it suffices to show that

$$\forall \mathbf{x}_0 \in \mathbb{R}^n \quad \mathbb{E}(V(\mathbf{x}(t))|\mathbf{x}(0) = \mathbf{x}_0) \leq V(\mathbf{x}_0)$$

By Dynkin's formula, one has for all $\mathbf{x}_0 \in \mathbb{R}^n$

$$\begin{array}{rcl} \mathbb{E}_{\mathbf{x}_0} V(\mathbf{x}(t)) & = & V(\mathbf{x}_0) + \mathbb{E}_{\mathbf{x}_0} \int_0^t \widetilde{A}V(\mathbf{x}(s))ds \\ & \leq & V(\mathbf{x}_0) - \lambda \mathbb{E}_{\mathbf{x}_0} \int_0^t V(\mathbf{x}(s))ds \leq V(\mathbf{x}_0) \end{array}$$

where $\mathbb{E}_{\mathbf{x}_0}(\cdot) = \mathbb{E}(\cdot|\mathbf{x}(0) = \mathbf{x}_0)$.

A.2 A variation of Gronwall's lemma

Lemma 4 *Let* $g : [0, \infty[\to \mathbb{R}$ *be a continuous function,* C *a real number and* λ *a* strictly positive *real number. Assume that*

$$\forall u, t \quad 0 \leq u \leq t \quad g(t) - g(u) \leq \int_u^t -\lambda g(s) + C ds \tag{A.2}$$

Then

$$\forall t \geq 0 \quad g(t) \leq \frac{C}{\lambda} + \left[g(0) - \frac{C}{\lambda}\right]^{+} e^{-\lambda t} \tag{A.3}$$

where $[\cdot]^{+} = \max(0, \cdot)$.

Proof **Case 1 :** $C = 0$, $g(0) > 0$.

Define $h(t)$ by

$$\forall t \geq 0 \quad h(t) = g(0)e^{-\lambda t}$$

Remark that h is positive with $h(0) = g(0)$, and satisfies (A.2) where the inequality has been replaced by an equality

$$\forall u, t \quad 0 \leq u \leq t \quad h(t) - h(u) = -\int_{u}^{t} \lambda h(s)ds$$

Consider now the set $S = \{t \geq 0 \mid g(t) > h(t)\}$. If $S = \emptyset$ then the lemma holds true. Assume by contradiction that $S \neq \emptyset$. In this case, let $m = \inf S < \infty$. By continuity of g and h and by the fact that $g(0) = h(0)$, one has $g(m) = h(m)$ and there exists $\epsilon > 0$ such that

$$\forall t \in]m, m+\epsilon[\quad g(t) > h(t) \tag{A.4}$$

Consider now $\phi(t) = g(m) - \lambda \int_{m}^{t} g(s)ds$. Equation (A.2) implies that

$$\forall t \geq m \quad g(t) \leq \phi(t)$$

In order to compare $\phi(t)$ and $h(t)$ for $t \in]m, m+\epsilon[$, let us differentiate the ratio $\phi(t)/h(t)$.

$$\left(\frac{\phi}{h}\right)' = \frac{\phi' h - h'\phi}{h^2} = \frac{-\lambda g h + \lambda h \phi}{h^2} = \frac{\lambda h(\phi - g)}{h^2} \geq 0$$

Thus $\phi(t)/h(t)$ is increasing for $t \in]m, m+\epsilon[$. Since $\phi(m)/h(m) = 1$, one can conclude that

$$\forall t \in]m, m+\epsilon[\quad \phi(t) \geq h(t)$$

which implies, by definition of ϕ and h, that

$$\forall t \in]m, m+\epsilon[\quad \int_{m}^{t} g(s)ds \leq \int_{m}^{t} h(s)ds \tag{A.5}$$

Choose now t_0 such that $m < t_0 < m + \epsilon$, then one has by (A.4)

$$\int_{m}^{t_0} g(s)ds > \int_{m}^{t_0} h(s)ds$$

which clearly contradicts (A.5).

Case 2 : $C = 0$, $g(0) \leq 0$

Consider the set $S = \{t \geq 0 \mid g(t) > 0\}$. If $S = \emptyset$ then the lemma holds true. Assume by contradiction that $S \neq \emptyset$. In this case, let $m = \inf S < \infty$. By continuity of g and by the fact that $g(0) \leq 0$, one has $g(m) = 0$ and there exists ϵ such that

$$\forall t \in]m, m+\epsilon[\quad g(t) > 0 \tag{A.6}$$

Let $t_0 \in]m, m+\epsilon[$. Equation (A.2) implies that

$$g(t_0) \leq -\lambda \int_m^{t_0} g(s)ds$$

which clearly contradicts (A.6).

Case 3 : $C \neq 0$

Define $\hat{g} = g - C/\lambda$. One has

$$\forall u,t \quad 0 \leq u \leq t \quad \hat{g}(t) - \hat{g}(u) = g(t) - g(u) \leq \int_u^t -\lambda g(s) + C ds = -\int_u^t \lambda \hat{g}(s)ds$$

Thus $\hat{g}$ satisfies the conditions of Case 1 or Case 2, and as a consequence

$$\forall t \geq 0 \quad \hat{g}(t) \leq [\hat{g}(0)]^+ e^{-\lambda t}$$

The conclusion of the lemma follows by replacing $\hat{g}$ by $g - C/\lambda$ in the above equation. □

References

[1] N. Aghannan and P. Rouchon. An intrinsic observer for a class of lagrangian systems. *IEEE Transactions on Automatic Control*, 48, 2003.

[2] D. Angeli. A lyapunov approach to incremental stability properties. *IEEE Transactions on Automatic Control*, 47:410–422, 2002.

[3] L. Arnold. *Stochastic Differential Equations : Theory and Applications.* Wiley, 1974.

[4] T. Caraballo and P. Kloeden. The persistence of synchronization under environmental noise. *Proceedings of the Royal Society A*, 461:2257–2267, 2005.

[5] T. Caraballo, P. Kloeden, and B. Schmalfuss. Exponentially stable stationary solutions for stochastic evolution equations and their pertubation. *Applied Mathematics and Optimization*, 50:183–207, 2004.

[6] L. d'Alto and M. Corless. Incremental quadratic stability. In *Proceedings of the IEEE Conference on Decision and Control*, 2005.

[7] B. Demidovich. Dissipativity of a nonlinear system of differential equations. *Ser. Mat. Mekh.*, 1961.

[8] H. Deng, M. Krstic, and R. Williams. Stabilization of stochastic nonlinear systems driven by noise of unknown covariance. *IEEE Transactions on Automatic Control*, 46, 2001.

[9] E. Dickmanns. *Dynamic Vision for Intelligent Vehicles*. Course Notes, MIT EECS dept, 1998.

[10] K. El Rifai and J.-J. Slotine. Contraction and incremental stability. Technical report, MIT NSL Report, 2006.

[11] W. Feller. *An Introduction to Probability Theory and Its Applications*. Wiley, 1968.

[12] P. Florchinger. Lyapunov-like techniques for stochastic stability. *SIAM Journal of Control and Optimization*, 33:1151–1169, 1995.

[13] P. Florchinger. Feedback stabilization of affine in the control stochastic differential systems by the control lyapunov function method. *SIAM Journal of Control and Optimization*, 35, 1997.

[14] V. Fromion. Some results on the behavior of lipschitz continuous systems. In *Proceedings of the European Control Conference*, 1997.

[15] B. Girard, N. Tabareau, Q.-C. Pham, A. Berthoz, and J.-J. Slotine. Where neuroscience and dynamic system theory meet autonomous robotics: a contracting basal ganglia model for action selection. *Neural Networks*, 2008.

[16] P. Hartmann. *Ordinary differential equations*. Wiley, 1964.

[17] R. Has'minskii. *Stochastic Stability of Differential Equations*. Sijthoff and Nordhoff, Rockville, 1980.

[18] D. Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43:525–546, 2001.

[19] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[20] J. Jouffroy. Some ancestors of contraction analysis. In *Proceedings of the IEEE Conference on Decision and Control*, 2005.

[21] J. Jouffroy and T. Fossen. On the combination of nonlinear contracting observers and uges controllers for output feedback. In *Proceedings of the IEEE Conference on Decision and Control*, 2004.

[22] H. Kushner. *Stochastic Stability and Control*. Academic Press, 1967.

[23] D. Lewis. Metric properties of differential equations. *American Journal of Mathematics*, 71:294–312, 1949.

[24] W. Lohmiller and J.-J. Slotine. On contraction analysis for nonlinear systems. *Automatica*, 34:671–682, 1998.

[25] W. Lohmiller and J.-J. Slotine. Nonlinear process control using contraction theory. *A.I.Ch.E. Journal*, 2000.

[26] W. Lohmiller and J.-J. Slotine. Contraction analysis of nonlinear distributed systems. *International Journal of Control*, 78, 2005.

[27] X. Mao. *Stability of Stochastic Differential Equations with Respect to Semimartingales*. Longman, White Plains, NY, 1991.

[28] Q.-C. Pham and J.-J. Slotine. Stable concurrent synchronization in dynamic system networks. *Neural Netw*, 20(1):62–77, Jan. 2007.

[29] V. Popov. *Hyperstability of Control Systems*. Springer-Verlag, 1973.

[30] W. Rudin. *Real and complex analysis*. McGraw-Hill, 1987.

[31] J.-J. Slotine and W. Li. *Applied Nonlinear Control*. Prentice-Hall, 1991.

[32] E. Sontag and Y. Wang. Output-to-state stability and detectability of nonlinear systems. *Systems and Control Letters*, 29:279–290, 1997.

[33] N. Tabareau and J.-J. Slotine. Notes on contraction theory. Technical report, MIT NSL Report, 2005.

[34] J. Tsinias. The concept of "exponential iss" for stochastic systems and applications to feedback stabilization. *Systems and Control Letters*, 36:221–229, 1999.

[35] W. Wang and J.-J. E. Slotine. On partial contraction analysis for coupled nonlinear oscillators. *Biol Cybern*, 92(1):38–53, Jan. 2005.

[36] Y. Zhao and J.-J. Slotine. Discrete nonlinear observers for inertial navigation. *Systems and Control Letters*, 54, 2005.